A Theory of Theories

by

P. A. Sturrock

July 1969

SUIPR Report No. 324

Prepared under

Air Force Office of Scientific Research
Contract F44620-69-C-0008
NASA Grant NGR 05 020 272
Office of Naval Research Contract
N00014-67-A-0112-0006

INSTITUTE FOR PLASMA RESEARCH
STANFORD UNIVERSITY, STANFORD, CALIFORNIA

36

A THEORY OF THEORIES

by

P. A. Sturrock

Air Force Office of Scientific Research
Contract F44620-69-C-0008
NASA Grant NGR-05-020-272
Office of Naval Research
Contract N00014-67-A-0112-0006

SUIPR Report No. 324

July 1969

Institute for Plasma Research
Stanford University
Stanford, California

ABSTRACT

The aim of this article is to set out a bookkeeping procedure by which a scientist (using the term flexibly) may compare the conclusions of a theory with facts obtained by reduction of observational data with the aim of assessing the hypothesis on which the theory is based. It is argued that the appropriate formalism is probability theory, and that the key process is the inductive process as represented by Bayes' theorem which indicates how the degree of belief in a hypothesis should be adjusted in response to new information.

The following model of the inductive process in science is adopted. Between observation and theory there is an "interface" which comprises a set of independent items: each item comprises a complete set of mutually exclusive statements. It must be possible to assign two probabilities to each of these statements, one by "reduction" of observational data, and the other by theoretical analysis of the considered hypothesis. The "observational" probabilities must be free from theoretical bias and vice versa. Formulas are derived which show (a) how the assumed probability of the hypothesis should be adjusted in response to information concerning one item, and (b) how such estimates concerning more than one item may be combined.

The model further requires that one should consider not one hypothesis but a complete set of mutually exclusive hypotheses. It is necessary to reconcile this requirement with the normal situation that a scientist has one or two specific theories to evaluate, the hypotheses of which do not form a complete set. A procedure is proposed to overcome this difficulty. One may compare a real analysis (or analyses) of a specified hypothesis (or hypotheses) with a "null" analysis of the complement of this hypothesis (or hypotheses). A "null" analysis is that which admits complete ignorance about the conclusions to be drawn from a hypothesis: the relevant probabilities may therefore be determined without specific knowledge of the hypothesis. If a specified theory fares worse than a "null" theory, it is a bad theory.

The method is illustrated by a "worksheet" indicating the way in which a few observational facts about pulsars may be used to appraise both the neutron-star hypothesis and the white-dwarf hypothesis.

CONTENTS

I. INTRODUCTION

This article is concerned with the role of induction in scientific research. However the principal aim is not to undertake a philosophical inquiry *per se*, but rather to set up a bookkeeping procedure for organizing the judgments involved in comparing a scientific theory with scientific data. For my own convenience, I shall draw examples from astrophysics, but I hope and believe that the ideas and methods could be useful in other fields also.

A monograph on quasars[1] by Kahn and Palmer gives an example of the type of judgment which scientists must attempt to make. What is unusual about the example is that the authors have the candor to present their judgment in numerical form. Table 3, on page 111 of that monograph, gives the "estimated probability of correctness" of six hypotheses concerning quasars. A similar appraisal was made by Professor L. Woltjer at the Conference on Seyfert Galaxies held at Tucson, Arizona in February 1968.

Although judgments of the type quoted above provide an effective means of communicating degrees of belief, they immediately raise a significant question: "How were these estimates arrived at?" In the examples quoted, there is no hint of the answer to this question.[2] An estimate of this type, without a description of the process by which it was made, invites controversy. The principal aim of this article is to present a procedure for arriving at estimates of this type. This does not guarantee that there will be no controversy, but it should help to localize the area of disagreement and so make the controversy more profitable.

The examples quoted have already established one relevant point: The appropriate formalism to use in investigating this problem is probability theory. Our aim then is to set up a model for the reasoning process involved in evaluating a scientific theory, and to analyze this model by the theory of probability. As I. J. Good[3] has remarked, "Probability is a part of reasoning and is therefore more fundamental than most theories." It is in this sense that the present article may be regarded as a "theory of theories."

Before proceeding, it is important to state that we shall not be concerned with a possible comparison between "perfect data" (observational or experimental) and a "perfect theory." Even if we knew what these

terms meant, their discussion would be irrelevant to everday life. Our aim is to determine how one can make a judgment about a theory which is admittedly uncertain and incomplete, in comparing it with data which are uncertain and incomplete. We further recognize that such an evaluation must be made not once, but progressively, as the data come in and as the theory develops. The following remarks of Jeffreys[4] are relevant here: "Either we can learn from experience or we cannot. The ability to learn from experience demands the concept of probability in relation to varying data, and the recognition of the meanings of more probable than and less probable than."

2. INDUCTION AND BAYES' THEOREM

Although textbooks frequently represent a science, such as physics, as being deductive, scientists are well aware that this is a characteristic of textbooks, not of science. P. G. Bergman, speaking at the Fourth Texas Conference on Relativistic Astrophysics in Dallas in 1968, stated "'Let the facts lead you where they may' is a gross oversimplification of how to proceed in science and not necessarily philosophically justified." The difference between deduction and induction has been pointed out very clearly by Polya[5] by means of the following examples.

The basic reasoning process of the deductive type is the syllogism:

A implies B.
B is false.
Therefore A is false.

By contrast, the inductive process follows a pattern such as the following:

A implies B.
B is true.
Therefore A is more credible.

The above example, although descriptive, does not lend itself to numerical evaluation. The basic procedure for making quantitative estimates of inductive arguments is provided by Bayes' theorem. According to Jeffreys,[6] "This theorem is to the theory of probability what Pythagoras's theorem is to geometry."

We introduce the notation $(A|B)$ to denote the probability that proposition A is true on the basis of the knowledge that proposition B is true. We adopt the convention that the measure of probability extends over the range 0 to 1: $(A|B) = 0$ if A is impossible given B; and $(A|B) = 1$ if A is certain given B.

The notation AB stands for the "product" of the two propositions A and B. Then AB is true if and only if both A and B are true. The "product rule" of probability theory[7] then states that

$$(AB|C) = (A|BC)(B|C) . \tag{2.1}$$

However, since AB = BA, this may alternatively be written as

$$(AB|C) = (B|AC)(A|C) . \tag{2.2}$$

We now see from these two equations that

$$(A|BC) = \frac{(B|AC)}{(B|C)} (A|C) . \tag{2.3}$$

This is Bayes' theorem.

In order to show the relationship of this theorem to the scientific method, we consider the following "model" of the scientific process. A certain hypothesis H is to be evaluated by comparing the theoretical consequences of this hypothesis with observation (either of the world as we find it or of a contrived situation called an "experiment"). Note here an important point: It must be possible to formulate a statement which makes sense <u>either</u> when it is regarded as the consequence of a theoretical analysis, <u>or</u> when it is regarded as the result of an observation. Since this statement forms a crucial "link" between theory and observation, we use for it the symbol L. We also introduce the symbol X to denote all information available to the scientist in addition to (and, we assume, preceding) his knowledge, derived from observation, that L is true. Then Bayes' theorem may be rewritten as

$$(H|LX) = \frac{(L|HX)}{(L|X)} (H|X) . \tag{2.4}$$

The term $(H|X)$ is the "prior probability" of the hypothesis H (prior, that is, to knowledge of L). The term $(H|LX)$ is the "post probability" of H, based on knowledge of both X and L. Bayes' theorem then tells us that the post probability equals the prior probability multiplied by a certain factor $(L|HX)/(L|X)$ which is sometimes termed the "likelihood." This is the ratio of the probability that L is true, given H and information X, to the probability that L is true, given only the information X.

Equation (2.4) gives us a procedure for adjusting our degree of belief in a hypothesis on the basis of incoming observational evidence. If H is irrelevant to L, then $(L|HX) = (L|X)$ so that $(H|LX) = (H|X)$: the probability is, appropriately enough, not affected by knowledge that L is true. If L seems quite likely on the basis of knowledge X, but very unlikely on the basis of X and the hypothesis H, then the probability that H is true is greatly reduced by the knowledge that L is true. If, on the other hand, L seems quite unlikely on the basis of information X alone, but quite likely on the basis of information X and the hypothesis H, then the fact that L is known (observationally) to be true greatly increases the probability that H is true.

It is worth noting a few further aspects of this theorem.

1. If $(L|X) = 0$, there is something wrong with our information X, since L is incompatible with X, yet L is observed to be true.

2. Assuming that $(L|X) \neq 0$, $(H|LX) = 0$ if $(H|X) = 0$. That is, an impossible hypothesis remains impossible, no matter what the evidence.

3. Denoting by $\bar{H}$ the negative of H, so that "$\bar{H}$ is true" means "H is not true", we see from (2) that $(\bar{H}|LX) = 0$ if $(\bar{H}|X) = 0$. This may be stated alternatively as $(H|LX) = 1$ if $(H|X) = 1$, since by the sum rule of probability theory,[7]

$$(\bar{H}|X) + (H|X) = 1, \quad \text{etc.} \tag{2.5}$$

In other words, if H is certain before we have knowledge of L, it will be certain after we have knowledge of L, no matter what L may be.

We see from the above paragraph that one should be very careful about assigning probability zero or unity to any proposition, since this entails that we can never change these estimates, no matter what subsequent infromation may turn up. I. J. Good has some useful advice on this point:[8] "Probability judgments can be sharpened by laying bets at suitable odds. If people always felt obliged to back their opinions when challenged, we would be spared a few of the 'certain' predictions that are so freely made." Whether or not we make wagers, we should, for the sake of future credibility, be very cautious about making "certain" theoretical predictions or stating "certain" observational facts: Theorists sometimes find a calculation to be wrong, and observers sometimes find that their results are not supported by subsequent observations by other groups.

3. MODEL OF THE INDUCTIVE PROCESS IN SCIENCE

Equation (2.4) of the preceding section, and the discussion which followed it, give a rough approximation to the inductive process in science. The simple model given in Section 2 would be adequate in a simple situation in which theory could predict the reading which one should obtain from a particular measuring device such as a meter. This however is not the usual situation in science. It may take a great deal of juggling and maneuvering to find a quantity which can be both measured and calculated.

Indeed, although one thinks of measurement as being the key process in exact sciences, many of the comparisons made between theory and observation are not normally expressed in terms of measurable quantities. For instance, one may require of a physical theory that it should be covariant under some transformation (Galilean, Lorentz, etc.). In the early studies of quasars, one of the most important questions was to determine whether the objects are "intragalactic" or "extragalactic." Similarly, the nature of pulsars for sometime hinged upon the question "Is a pulsar a white dwarf or a neutron star?"

Our first aim in this section is to set up a "model" for the inductive process in science which is a closer approximation to the methods actually used than the simple model presented in Section 2. The essential requirement is to be able to make a comparison between theory and observation. Hence a key requirement is for an "interface" between theory and observation. It seems that the basic requirement for such an interface is that there should be a language which can be understood both by theorists and by observers. More specifically, we adopt the following definition of the interface for the purpose of constructing a model: The interface comprises a number of statements, each of which is both (a) a possible result of data reduction of observations, and (b) a possible consequence of theoretical analysis of a hypothesis under consideration.

It is convenient to make further assumptions about these statements. We assume that they may be arranged into groups: Each group of statements will be termed an "item." We assume that there is a finite set of items I_α, $\alpha = 1, 2, \ldots, A$.

With each item I_α there is to be associated a group of statements which, for present convenience, we assume to be finite in number. (This assumption can be relaxed without difficulty and must be relaxed when dealing with statements about continuous variables.) This set of statements will be represented by $S_{\alpha n}$, $n = 1, 2, \ldots, N_\alpha$. For convenience, we assume that the group of statements associated with any item are mutually exclusive and complete. That is, for any item such as I_α, one and only one of the statements $S_{\alpha n}$ is tr e.

When a science is well established, one tends to forget that the theory of that science is simply a construct, the validity of which is to be established by comparison with observations. Indeed, one can easily come to regard the abstract construct as having a "reality" of its own. A remark by Eddington[9] is particularly appropriate here: "Physical science may be defined as 'the systematization of knowledge obtained by measurement.' It is a convention that this knowledge shall be formulated as a description of a world--called the 'physical universe'." It is necessary for us to distinguish between observational data and theoretical data, and to consider explicitly the connection between them.

We first consider what goes on on the observational side of the interface. An astronomer obtains his information by means of photographs, spectra, radio records, etc., whichever happens to be the output of the observational instrument he is using. An experimenter collects similar "raw data." However, neither observer nor experimenter transmits this raw data to his scientific colleagues. The transaction between the observer (this term being used to include "experimenter") and the theorist is typically the publication of an article in which the observational results are presented and analyzed. Sometimes the conclusions which the observer draws from his data have almost the certainty of mathematical deduction. Usually, however, there are a number of assumptions and provisos which are explicitly or implicitly involved in going from the data to the conclusions. Furthermore, this process of "data reduction" probably involves the use of theoretical knowledge. One must pay careful attention to this aspect of data reduction, if one's aim is to compare the results of the observation with one or more theories.

The basic rule of data reduction is that, if theoretical knowledge is to be used, it should be carefully specified, and preferably should comprise only theoretical knowledge which is beyond dispute. If the aim of an observation is to obtain information with which to evaluate one or more theoretical hypothesis, the data reduction must studiously avoid any steps which explicitly or implicitly appeal to these hypotheses.

Data reduction can range from a very simple to a very sophisticated operation. A new technique in data reduction may represent an important step forward in a science. The construction of the Hertzsprung-Russell diagram[10,11] is a case in point. In looking at raw data, one may not be able to see the wood for the trees. The aim of good data reduction is to enable one to see the shape of the forest. A Hertzsprung-Russell diagram might be prepared from many precise observations, yet the significant information in the diagram may comprise the approximate clustering of points to form a simple and not-to-well-defined curve.

In the present model, the result of data analysis is to provide varying degrees of support for alternative statements $S_{\alpha m}$ of each item I_{α}, on the basis of observational knowledge which we denote by O. In this model, the sum total of observational knowledge, as it may be used

for comparison with theory, is provided by the set of probabilities $(S_{\alpha n}|ROX)$. R denotes the process of data reduction. This symbol should be introduced since different reduction procedures may lead to different estimates of the reduced data.

Since, for each α, the set of statements $S_{\alpha n}$ are complete and mutually exclusive, we see from the sum rule of probability theory[7] that

$$\sum_{n=1}^{N_\alpha} (S_{\alpha n}|X) = 1, \quad \sum_{n=1}^{N_\alpha} (S_{\alpha n}|ROX) = 1, \text{ etc.} \tag{3.1}$$

In the present context, X denotes all information which is not in question, including theoretical knowledge used in the data reduction and, conversely, whatever observational knowledge the theorist is to be permitted to use in his analyses.

We now look at the other side of the fence, and inquire about operations on the theoretical side of the interface. The basic procedure has been described by Jeffreys[12], in his discussion of 'theory": "The use of the word 'theory' in several different senses is perhaps responsible for a good deal of confusion. What I prefer to call an 'explanation' consists of several parts: First, a statement of a hypothesis; secondly, the systematic development of its consequences; thirdly, the comparison of these consequences with observation."

However, Jeffreys elsewhere[13] makes the following relevant and qualifying remarks: "We get no evidence for a hypothesis by merely working out its consequences and showing that they agree with some observations, because it may happen that a wide range of other hypotheses would agree with those observations equally well. To get evidence for it we must also examine its various contradictories and show that they do not fit the observations. This elementary principle is often overlooked in alleged scientific work, which proceeds by stating a hypothesis, quoting masses of results of observation that might be expected on the hypothesis and possibly on several contradictory ones, ignoring all that would not be expected on it, but might be expected on some alternative, and claiming that the observations support the hypothesis. . . . So long as

alternatives are not examined and compared with the whole of the relevant data, a hypothesis can never be more than a considered one."

We see from the above that the principal job of a theorist is to determine by theoretical analysis the consequences of one or more hypotheses. As Jeffreys points out, one obtains the strongest evidence for a hypothesis when one analyzes, and compares with observations, this hypothesis and also whatever additional hypotheses are necessary to form a complete set. An important case for us to consider is, therefore, that we are considering a set of hypotheses H_i, $i = 1, \ldots, I$, which are mutually exclusive and form a complete set. We assume that each of these hypotheses is subjected in turn to an analysis A so that one arrives at probabilities for the statements $S_{\alpha n}$, where these statements are now regarded as consequences of each hypothesis considered in turn. In this way we would arrive at the probabilities $(S_{\alpha n}|AH_iX)$. It may well be that, if the problem is sufficiently well defined, each hypothesis H_i implies a definite statement $S_{\alpha n}$ of each item I_α, so that each of the probabilities $(S_{\alpha n}|AH_iX)$ would be either unity or zero. However, there may be some uncertainty in the basic information X, and it is normally the case that scientific analysis is incomplete and imperfect. Hence we should expect that, in general, the probabilities $(S_{\alpha n}|AH_iX)$ will have values between zero and unity.

Although it is most desirable to consider any hypothesis under consideration as a member of a complete set, and to analyze all members of that set, this will often be impractical in normal scientific theoretical work. It may be that the hypotheses can be identified, but are too many in number for each of them to be analyzed. Another possibility is that it is very difficult, or practically impossible, to identify hypotheses which build up a given hypothesis into a complete set. It is therefore important for us to have some way of proceeding which does not hinge upon the explicit identification and analysis of a complete set of hypotheses.

Jeffreys recognizes that a well established hypothesis will be accepted simply on the basis of the agreement between consequences of that hypothesis and observation. He states[14] "The chief advances in modern physics . . . were achieved by the method of Euclid and Newton: to state a set of hypotheses, work out their consequences, and assert them if they

accounted for most of the outstanding variation." It is therefore necessary for us to specify a procedure which involves the detailed specification and analysis of one or more hypotheses $H, \ldots, H_I$, in the case that this set of hypotheses does <u>not</u> form a complete set. We assume, however, that the hypotheses are chosen to be mutually exclusive.

We assume that it is possible to expand this set of hypotheses to form a complete set by adding a hypothesis H_0 which excludes and is excluded by any one of the hypothesis H_i, $i = 1, \ldots, I$. (If the set were to be specified explicitly, it might be more convenient to express H_0 as the disjunction of a large number of hypotheses, but for present purposes this consideration is irrelevant.) Thus the set of hypotheses $H_0, H_1, \ldots H_I$ are mutually exclusive and form a complete set.

We now ask what information we can obtain about the hypothesis H_0 without specifying the hypothesis or carrying out an analysis of the hypothesis. The answer is, of course, that in this circumstance we must remain ignorant about H_0. However, this does not mean that we cannot fit H_0 into our model of the inductive process in science. We have recognized that theoretical analysis is in practice imperfect. We can introduce, as an extreme case, a "null" analysis A_0 which gives no information whatever about the consequences of the hypothesis to which it is applied. It is then possible to maintain the formalism which is based on the assumption that we are analyzing the consequences of a complete set of hypotheses, by adopting the following strategem: first, we assume that the known hypotheses $H_1, \ldots, H_1$, are supplemented by a hypothesis H_0 to form a complete set; second, we assume that, whereas each hypothesis $H_1, \ldots, H_I$ is subjected to a proper analysis A, the hypothesis H_0 (which is not to be specified explicitly) is assumed to be subjected to the "null" analysis A_0. We shall, for simplicity of notation, suppress the symbols A and A_0 in the remainder of this section.

Now suppose that we chose to identify $(S_{cm}, H_0 X)$ with (S_{cm}, X). Since the hypotheses $H_0, H_1, \ldots, H_I$ form a complete set, we know that $H_0 + H_1 + \ldots + H_I$ is true, where the summation sign here indicates a "logical sum," i.e., "and/or." Hence

$$\left(S_{\alpha n} \sum_{i=0}^{I} H_i \middle| X\right) = (S_{\alpha n}|X) \quad . \tag{3.2}$$

On noting that the hypotheses $H_0, H_1, \ldots, H_I$ are mutually exclusive, we see from the sum rule of probability theory[7] that

$$\left(S_{\alpha n} \sum_{i=0}^{I} H_i \middle| X\right) = \sum_{i=0}^{I} (S_{\alpha n} H_i|X) \quad . \tag{3.3}$$

We may now use the produce rule[7] to write

$$(S_{\alpha n} H_i|X) = (S_{\alpha n}|H_i X)(H_i|X) \quad , \tag{3.4}$$

so that

$$(S_{\alpha n}|X) = \sum_{i=0}^{I} (S_{\alpha n}|H_i X)(H_i|X) \quad . \tag{3.5}$$

If we now make the choice

$$(S_{\alpha n}|H_0 X) = (S_{\alpha n}|X) \quad , \tag{3.6}$$

we see that

$$(S_{\alpha n}|H_0 X) = [1 - (H_0|X)]^{-1} \sum_{i=1}^{I} (S_{\alpha n}|H_i X)(H_i|X) \quad . \tag{3.7}$$

It follows from this equation that if a particular statement $S_{\alpha' n'}$ is impossible on the basis of hypotheses $H_1, \ldots, H_I$, then it must be considered to be impossible on the basis of hypothesis H_0 also. This represents a defect of the convention for $(S_{\alpha n}|H_0 X)$ specified by Equation (3.6). We would like the probabilities $(S_{\alpha n}|H_0 X)$ to be "maximally noncommittal" about the statements $S_{\alpha n}$, subject only to restrictions

imposed by the information X. The manner in which one may ascribe "maximally noncommittal" values to a set of statements, taking account of possible information about the statements, has been discussed by Jaynes.[15] We will not pursue this point here. For our purpose, it is important only to note that it is necessary to specify the probabilities $(S_{\alpha m}|H_0)$ in as noncommittal a manner as possible, and that it is not desirable to adopt the convention (3.6).

Although we have recognized two very different theoretical situations, it has now been possible to set up a formalism which covers both cases. The formulas which we shall derive in the next two sections may be applied to either case. The difference between the two cases will become significant only when we consider the information which is to be fed into the formulas, or the interpretation to be placed upon estimates made by means of the formulas. We shall discuss this further in the final section.

4. INDUCTION USING ONE FACT

We now suppose that, by data reduction, the observer has assigned various probabilities to the statements which constitute the interface. The theorist has assigned probabilities to the same statements based on analysis of each considered hypothesis, and also possibly based on a null analysis of a supplemental hypothesis. This information, and knowledge of the prior probability of each hypothesis, should enable us to assign a post probability to each hypothesis. That is, given the information specified in the preceding section, and given the values of $(H_i|X)$, we should be able to calculate $(H_i|OX)$. Here and in subsequent sections it is convenient to suppress the symbols R, A and A_0. These are implicitly present: the observations have been reduced by a specified procedure; hypotheses $H_1, \ldots, H_I$ have been reduced by analysis A and hypothesis H_0 (if it must be introduced) has been reduced by the null analysis A_0. In line with this change of notation, we shall refer to H_0 as the "ignorance hypothesis."

In this section, we make the simplifying assumption that there is only one item to be considered. If we regard the observational knowledge pertaining to an item as a "fact," this means that we have only one fact to consider. In this section, therefore, we may drop the suffix α.

As an aside, we may note that this can be regarded as a formal change rather than a substantive change, since it is always possible to combine items. Specifically, we could introduce the notation

$$S_{nn'n''\dots} = S_{1n}\, S_{2n'}\, S_{3n''} \dots \tag{4.1}$$

The set of statements $S_{nn'n''}$ again are mutually exclusive and form a complete set, and they comprise all the information represented by the separate groups of statements. However, our intention is to consider one item in this section, and to consider in the next section how one should combine information derived from several separate items.

Our aim is to calculate $(H_i|OX)$, the "post probabilities" of the various hypotheses as determined by the prior information X and the observations O. Since the statements S_n are mutually exclusive and form a complete set, this probability may be written as

$$(H_i|OX) = \left(H_i \sum_n S_n \middle| OX\right) , \tag{4.2}$$

where the summation sign indicates a "logical summation." By the sum rule of probability theory,[7] this equation may be expressed as

$$(H_i|OX) = \sum_n (H_i S_n|OX) , \tag{4.3}$$

and use of the product rule[7] enables us to put this equation in the following form:

$$(H_i|OX) = \sum_n (H_i|S_n OX)(S_n|OX) . \tag{4.4}$$

According to the rules of our model, the connection between the hypotheses and the observations occurs only via the statements S_n. If it

is asserted that S_n is true, all other knowledge about the observations is irrelevant, as far as the hypotheses are concerned. This property of our model therefore implies that

$$(H_i|S_n OX) = (H_i|S_n X) \quad , \tag{4.5}$$

in consequence of which Equation (4.4) becomes

$$(H_i|OX) = \sum_n (H_i|S_n X)(S_n|OX) \quad . \tag{4.6}$$

At this stage we use the following form of Bayes' theorem [Equation (2.3)];

$$(H_i|S_n X) = \frac{(S_n|H_i X)}{(S_n|X)} (H_i|X) \quad . \tag{4.7}$$

This introduces the probabilities $(S_n|X)$, which do not appear among our initial data. It is at this point that we profit from the assumption that the set of hypotheses H_i are mutually exclusive and form a complete set. We saw in Section 3 that these assumptions lead to Equation (3.5), which is now written as

$$(S_n|X) = \sum_j (S_n|H_j X)(H_j|X) \quad , \tag{4.8}$$

leaving the range of summation $(0, 1, \ldots, N$ or $1, 2, \ldots, N)$, unspecified. Then Equation (4.7) becomes

$$(H_i|S_n X) = \frac{(S_n|H_i X)}{\sum_j (S_n|H_j X)(H_j|X)} (H_i|X) \quad . \tag{4.9}$$

On combining Equations (4.4) and (4.9), we finally arrive at

$$(H_i|OX) = \sum_n \left[\frac{(S_n|H_iX)(S_n|OX)}{\sum_j (S_n|H_jX)(H_j|X)} \right] (H_i|X) \quad . \tag{4.10}$$

This equation is the principal result of this section. The above formula for the post probabilities $(H_i|OX)$ involves only the prior probabilities $(H_i|X)$ and the probabilities of statements S_n as determined on the one hand by reduction of the observations O and on the other hand by analysis of the hypotheses H_i. We may note the following desirable property of this equation

$$\sum_i (H_i|OX) = 1 \quad . \tag{4.11}$$

This shows that the formula (4.10) will never yield a probability of a hypothesis greater than unity, and that it will show the probability of one hypothesis to be equal to unity only if it shows the probabilities of all other hypotheses to be zero.

5. INDUCTION USING MANY FACTS

In the preceding section we obtained a formula to describe the probabilities to be assigned to a set of hypotheses on the basis of the prior probabilities and of observational and theoretical knowledge concerning one item of the interface. We now consider how we may take account of knowledge concerning several such items. We assume that these items are "independent" in a sense to be specified later.

It is convenient to introduce F_α as the "fact" associated with the item I_α. According to our model, the fact F_α comprises the set of probabilities $(S_{\alpha n}|OX)$, $n = 1, \ldots, N_\alpha$.

We now suppose that a group of hypotheses H_i have been evaluated in terms of two facts F_1 and F_2, considered separately and independently. In this way we have arrived at probabilities which may be written

as $(H_i|F_1X)$, $(H_i|F_2X)$. The problem which we now consider is that of determining the probabilities $(H_i|F_1F_2X)$. The sense in which F_1 and F_2 are considered to be independent is the following: knowledge of F_1 will influence our interpretation of F_2 only through the effect which F_1 has on our evaluation of the hypotheses H_i and the influence of knowledge of hypotheses H_i on our interpretation of F_2. We assume of course that the converse also is true.

We first note that $(H_i|F_1F_2X)$ may be expressed as follows:

$$(H_i|F_1F_2X) = \frac{(H_iF_1|F_2X)}{(F_1|F_2X)} \quad . \tag{5.1}$$

By an argument parallel to that leading to Equation (4.4), we see that

$$(H_iF_1|F_2X) = \sum_j (H_iF_1|H_jF_2X)(H_j|F_2X) \tag{5.2}$$

and

$$(F_1|F_2X) = \sum_j (F_1|H_jF_2X)(H_j|F_2X) \quad . \tag{5.3}$$

We now note that the first term on the right-hand side of (5.2) may be expressed as

$$(H_iF_1|H_jF_2X) = (H_i|F_1H_jF_2X)(F_1|H_jF_2X) \quad . \tag{5.4}$$

However, since the set of hypotheses is assumed to be mutually exclusive and complete,

$$(H_i|F_1H_jF_2X) = \delta_{ij} \quad . \tag{5.5}$$

Furthermore, our specification of the sense in which F_1 and F_2 are taken to be independent implies that

$$(F_1 | H_j F_2 X) = (F_1 | H_j X) \quad . \tag{5.6}$$

If we now note from Bayes' theorem that

$$(F_2 | H_j X) = \frac{(H_j | F_2 X)}{(H_j | X)} (F_2 | X) \quad , \tag{5.7}$$

we find from Equations (5.2) through (5.7) that Equation (5.1) may be expressed as

$$(H_i | F_1 F_2 X) = \frac{(H_i | F_1 X)(H_i | F_2 X)[(H_i | X)]^{-1}}{\sum_j (H_j | F_1 X)(H_j | F_2 X)[(H_j | X)]^{-1}} \quad . \tag{5.8}$$

It is a straightforward matter to prove (by induction!) that the general formula is

$$(H_i | F_1 \ldots F_A X) = \frac{(H_i | F_1 X) \ldots (H_i | F_A X)[(H_i | X)]^{-(A-1)}}{\sum_j (H_j | F_1 X) \ldots (H_j | F_A X)[(H_j | X)]^{-(A-1)}} \quad . \tag{5.9}$$

We note from this equation that

$$\sum_j (H_i | F_1 \ldots F_A X) = 1 \quad . \tag{5.10}$$

6. WORK-SHEET CONCERNING THE THEORY OF PULSARS

Since the principal aim of this article is a bookkeeping procedure rather than a philosophical inquiry, it seems expedient to present a simple example of the use of the formulas which we have derived, before indulging in a philosophical discussion of the their significance.

The example to be discussed concerns the current astrophysical problem of the nature of pulsars.[16] Although it is now generally agreed that

a pulsar is to be identified with a rotating neutron star, less than a year ago there was a lively controversy concerning two possibilities: the rotating neutron star model and the pulsating white dwarf model. In January 1969 the evidence had become strongly favorable to the neutron star hypothesis, and I was interested to try to assess the strength of the case by using the techniques presented in this article. The following material is therefore to be regarded as a personal worksheet (which is now several months out of date), not a valid description of the present state of knowledge concerning pulsars.

The "work-sheet" is presented in Table 1. Four items are here considered.

1. The range of period.
2. The change of period.
3. Connection with supernovae.
4. Absence of photospheric radiation.

These items will be discussed in turn. In addition to the hypotheses that the object is a neutron star (H_1) and that the object is a white-dwarf (H_2), we consider the "ignorance" hypothesis (H_0), allowing for the possibility that there may be some other hitherto unsuspected explanation.

Before discussing the above items, we should recognize that one item is conspicuous by its absence: namely, radio emission. The reason this was not considered is that very little persuasive information was available concerning radio emission from a neutron-star model or from a white-dwarf model. Hence, in this respect, each model would fare no better and no worse than the ignorance hypothesis. Such an item need not be considered explicitly - it is "ignorable." Similarly, if there is no observational information about an item, it is ignorable.

There is another interesting aspect to the (absent) item concerned with radio emission. It is concerned not with a possible property of pulsars, but with a necessary property. In order for an object to be accepted as a pulsar, it must have certain properties. In the early days of a phenomenon, such properties will typically be observational. In order to clarify this point, we introduce for this purpose only a "zeroth" item:

Table 1. PULSAR WORK-SHEET

Based on prior probabilities $(H_0|X) = (H_1|X) = (H_2|X) = 1/3$

	Observation	Ignorance Hypothesis	Neutron-Star Hypothesis	White-Dwarf Hypothesis
Item 1. Range of Period				
S_{11}: Periods extend over range -.03 sec to 3 sec	1	.5	1	.01
S_{12}: S_{11} not true	0	.5	0	.99
Post probabilities		.331	.662	.007
Item 2: Change of Period				
S_{21}: All pulsars slow down	.97	.33	1	.09
S_{22}: Neither S_{21} nor S_{23} true	.03	.33	0	.01
S_{23}: All pulsars speed up	0	.34	0	.9
Post probabilities		.256	.681	.063
Item 3. Association with Supernovae				
S_{31}: All pulsars related to supernovae	.9	.03	.99	0
S_{32}: Some pulsars related to supernovae	.1	.01	.01	.01
S_{33}: No pulsars related to supernovae	0	.96	0	.99
Post probabilities		.060	.907	.033
Item 4. Photospheric Radiation				
S_{41}: No pulsars have detectable photospheric radiation	.99	.33	1	.01
S_{42}: Some pulsars have detectable photospheric radiation	.01	.33	0	.09
S_{43}: All pulsars have detectable photospheric radiation	0	.34	0	.9
Post probabilities		.257	.257	.010
Post probabilities based on all facts		.012	.988	1.37×10^{-8}

Item 0. Periodic Pulsed Radio Emission.

The statements might be chosen as follows:

S_{01}: The object emits detectable periodic pulsed radio emission.

S_{02}: The object does not emit detectable periodic pulsed radio emission.

Then a pulsar is "defined" by the statement S_{01}. Hence it becomes a convention, associated with the name "pulsar," that $(S_{01}|OX) = 1$. Note that the strict separation between observational facts and theoretical conclusions applies here also: the convention that $(S_{01}|OX) = 1$ should not prejudice the evaluation of $(S_{01}|HX)$ for any hypothesis.

We may also note that the definition of a phenomenon may at one time be observational in nature and at another time theoretical in nature. It may be, for instance, that when pulsars are "fully" understood, they will be defined by a hypothetical model[17].

We now return to a discussion of Items 1 - 4.

Item 1. The Range of Periods.

We adopt the following statements for this item.

S_{11}: The periods of pulsars extend over a range extending from .03 sec or below to 3 sec or above.

S_{12}: The range of periods does not extend as low as .03 sec and/or does not extend as high as 3 sec.

We know from observational evidence that statement S_{11} is correct. Hence we set $(S_{11}|OX) = 1$ and $(S_{12}|OX) = 0$.

Here and in Items 2 and 4, we assign equal probabilities to possible statements of an item, when regarded as consequences of the ignorance theory. Hence we assign $(S_{11}|H_0X) = .5$, $(S_{12}|H_0X) = .5$.

The neutron star hypothesis is compatible with any periods down to a minimum value of about 1 m sec. Hence we set $(S_{11}|H_1X) = 1$, $(S_{12}|H_1X) = 0$.

If the pulses were produced by pulsation of a white dwarf, one would expect the periods to extend over a range from about one second to several seconds. However, one cannot be completely confident that shorter periods

are quite impossible. Therefore, to be conservative, we say $(S_{11}|H_2X) = .01$, $(S_{12}|H_2X) = .99$.

We may now use Equation (4.10) to evaluate the post-probabilities of the three alternative hypotheses, using the information of Item 1. On denoting by F_1 the observational data relevant to Item 1, we find that $(H_0|F_1X) = .331$, $(H_1|F_1X) = .662$ and $(H_2|F_1X) = .007$.

Item 2. Rate of Change of Period.

The statements are as shown in the table. Five pulsars are known to be slowing down. No observable change has been detected for other pulsars, which we interpret to mean that the rate of change is too small for its sign to be determined. If we assign equal prior probabilities to the three possible statements, the fact that five pulsars are known to be slowing down leads to the observational probabilities shown in the table.

If pulsars are rotating neutron stars, we expect pulsars to slow down.

If pulsars are pulsating white dwarfs, we expect them to speed up, since the white dwarf should become smaller and denser as it ages. However, without specifying what kind of pulsation we are considering, we should assign a small probability to the possiblity that the pulsations will speed up. To be cautious, we should also make some allowance for the possibility that some pulsars slow down and some pulsars speed up. These considerations lead to the probabilities shown.

The post-probabilities based on data for this item are also shown.

Item 3. Relationship of Pulsars to Supernovae.

It is known that two pulsars are located in the same positions as supernova remnants. This suggests that all pulsars are related to supernovae (Statement S_{31}). However, one should allow for the possibility that only some pulsars are related to supernovae (Statement S_{32}). We can rule out the possibility that no pulsars are related to supernovae (Statement S_{33}).

The only idea which has been advanced concerning the creation of a neutron star is that it is formed during a supernova explosion. The

neutron-star hypothesis therefore leads us to expect that all pulsars should be related to supernovae. However, to be cautious, we might allow for the possibility that there is some other way in which neutron stars can be formed, and therefore assign a small probability, on the basis of H_1, to S_{32}.

The current view about white dwarfs is that they represent a state of senility reached by low-mass stars in a non-catastrophic evolution. However, the idea has been suggested that a supernova explosion may leave a white dwarf as end product. We therefore choose the probabilities shown.

In assigning probabilities to the statements of Item 3 for the ignorance hypothesis, the simplest procedure of assigning equal weight to the alternatives seems unacceptable, since a supernova remnant is an unusual object. The key question is: "What is the probability that objects which are (or appear to be) supernova remnants are associated with some unspecified object which is neither a neutron star nor a white dwarf?" Most astronomers would regard the probability as small, but we are alert to the fact that we must not set the probability as zero. Taking this question in isolation (i.e., neglecting all other evidence related to pulsars), I should regard .1 as too high and .01 as too low; accordingly, I choose $(S_{31}|H_0X) = .03$. I assign a somewhat lower probability to the possibility that, if pulsars are neither neutron stars nor white dwarfs, some of them happen to be related to supernova remnants or objects which look like supernova remnants, setting $(S_{32}|H_0X) = .01$. Hence $(S_{33}|H_0X) = .96$.

Item 4. Photospheric Radiation.

In only one case has a star-like optical object been identified with a pulsar: this is the south-preceding star of the Crab Nebula, identified with the Crab pulsar. However, it has been shown that the radiation of this "star" consists of pulses similar to the radio pulses, so that it cannot be interpreted as photospheric radiation. There is one normal star near the location of the pulsar CP 1919, but this is generally thought not to be the optical counterpart of the pulsar. To avoid being dogmatic, we choose the probabilities listed in the table.

A neutron star is too small for its photospheric radiation to be optically detectable. On the other hand, a white dwarf should be clearly visible at the typical distance estimated for a pulsar. However, the luminosity of a white dwarf decreases steadily as it gets older, so that it is possible that some white dwarfs of a sample would be invisible, and we cannot rule out the possibility that all white dwarfs of a given class may be invisible. These considerations are reflected in the probabilities listed in item 4.

We see from the Table that each fact counts against the white-dwarf hypothesis. However, none could be considered to be conclusive. In Items 1, 2 and 3, the neutron star hypothesis fares better than the ignorance hypothesis, but not much better. This is due simply to the fact that each item has been divided into only a small number of possible statements. In order to get strong evidence for a hypothesis in comparison with the ignorance hypothesis, it is necessary to consider an item divided into a large number of statements, or to have some reason for assigning very nonuniform weighting to the possible statements on the basis of the ignorance hypothesis, as in the case of Item 3.

The post-probabilities calculated from each of the four facts may be combined by means of Equation (5.9). When this is done, we arrive at the result

$$(H_0|OX) = .012$$

$$(H_1|OX) = .988$$

$$(H_2|OX) = 1.37 \times 10^{-6}$$

We see that the combined facts give very strong evidence against the white-dwarf hypothesis. If one were to consider the white dwarf hypothesis and the neutron star hypothesis as being the only possibilities, then the considerations listed above would show conclusively that pulsars are to be interpreted as neutron stars. If, however, we give equal prior-probability to the hypothesis that there is some other explanation, then we find that the evidence for the neutron-star hypothesis is good, but not overwhelming.

When probabilities become very close to unity or very close to zero, it is convenient to introduce a change of notation which gives a better feeling for the magnitude of the effect. The change of notation fits most easily with the "odds" notation where the odds on a proposition is defined as $P/(1-P)$, that is, the ratio of the probability that the proposition is true to the probability that it is not true. The odds on a proposition can clearly vary from zero to infinity, so that it is convenient to use logarithmic notation. Good[18] recommends that one use the decibel notation. McCamy[19], in a recent article, recommends the use of "brigg" (decibrigg, etc.) as a general term, in place of bel. However, the symbol "db" may be used for either "decibel" or "decibrigg," according to taste.

Using the symbol $\bar{H}$ to denote the proposition "H is not true," we may now express the above result as follows.

$$(H_0|OX)/(\bar{H}_0|OX) = .012$$

$$(H_1|OX)/(\bar{H}_1|OX) = 83$$

$$(H_2|OX)/(\bar{H}_2|OX) = 1.37 \times 10^{-6}$$

These results may be expressed in db-notation as follows:

Odds on H_0 = -19.2 db,

Odds on H_1 = 19.2 db,

Odds on H_2 = -58.6 db.

We see that, if the only admitted possibilities were the neutron-star hypothesis and the white-dwarf hypothesis, the odds on the neutron-star hypothesis would be about 87 db. However, if one admits the ignorance hypothesis, and gives it equal prior probability with the other two hypotheses, then the odds on the neutron-star hypothesis is only 15 db. This demonstrates a view of scientific theories which sometimes finds expression: it is easier to prove a theory wrong than to prove it right[20]. The high value of the odds on H_1, when H_0 is ignored, is really due to the high odds against H_2.

The above example shows that a combination of three or four quite cautious statements can lead to a strong result. We note also that, with one exception, the statements do not involve numbers. Hence the application of this type of formalism is not restricted to problems involving numerical data. It could be applied equally well to biology, criminology or social science.

We may also point out two defects of the above worksheet. A large value (.9) was assigned to $(S_{31}|OX)$. This was due to the fact that both the Crab pulsar and the Vela pulsar are associated with supernova remnants. However, this represents only two pulsars out of about twenty-six. If all pulsars should be equally likely to show this association, then the evidence is not too impressive. However, observable supernova remnants are only a few hundred or a few thousands years old so that this association should be observable only for young pulsars. If one adopts the neutron-star hypothesis, then one would expect to observe an association with supernova remnants only for short-period pulsars. This is what is found to be the case. Hence, in assigning a large value to $(S_{31}|OX)$, I was in fact making use of the hypothesis H_1. This means that I broke one of the basic rules of the game. The interface should therefore be adjusted accordingly, for instance by taking S_{31} to be the statement "short period pulsars are related to young supernova remnants."

At this date (June 1969), one would face a further difficulty in drawing up the above worksheet. It has been observed[21] that the Vela pulsar speeded up between 24 February and 3 March 1969, and then slowed down again. Should one therefore assert that $(S_{21}|OX) = 0$ and $(S_{22}|OX) = 1$? Strictly speaking, one should. But this would be misleading, in the sense that it would not represent the way science is done.

At this stage, the scientist would begin to modify what he regards as the "pulsar problem." He would say that the "basic" problem, representing the "normal behavior" of pulsars, is such that all pulsars slow down.[22] He would regard the behavior of the Vela pulsar as an anomaly, representing a minor secondary phenomenon, which he will probably not bother about until he has arrived at an adequate understanding of the "normal behavior." With this modification in the definition of the problem, one could still assert in June 1969 that $(S_{21}|OX) = .97$, $(S_{22}|OX) = .03$, $(S_{23}|OX) = 0$.

7. DISCUSSION

It is hoped that the theory which has been developed in the preceding sections will be interesting for two reasons: first, because it provides a way to combine observational and theoretical evidence to assess how well a theory explains an observational phenomenon; and, second, because one can learn some useful lessons from the exercise of trying to establish a procedure for making this assessment.

We saw in the previous section how the formalism may be used in practice. However, the formulas which we have derived and the example which we gave are such that each item is divided up into a finite number of alternative statements. In some cases, for instance when considering continuous measurable quantities, it will be necessary to consider a continuous sequence of statements. The required modification of Equation (4.10) is straightforward.

We now suppose that the statements are enumerated by a continuous variable ν rather than the discrete variable n. For instance, the statement S_ν may be the statement that the measurable quantity $\mathfrak{F}$ has the value $F(\nu)$. If we now denote by "S_ν to $S_{\nu+d\nu}$" the logical sum of all statements enumerated by ν as it runs from ν to $\nu+d\nu$, we can introduce the notation

$$(S_\nu \text{ to } S_{\nu+d\nu} | H_i X) = (S_\nu | H_i X)_\nu \, d\nu, \quad \text{etc.} \tag{7.1}$$

Using this notation, and replacing the summation sign in Equation (4.10) by an integration sign, we obtain the formula

$$(H_i | OX) = \left[\int d\nu \, \frac{(S_\nu | H_i X)_\nu \, (S_\nu | OX)_\nu}{\sum\limits_j (S_\nu | H_j X)_\nu \, (H_j | X)} \right] (H_i | X) \quad . \tag{7.2}$$

We now turn to some of the implications of the model. The first is that it is essential to set up an appropriate and meaningful interface between observational data and theoretical calculation. Getting to the interface from observations needs data reduction. An excellent example

of data reduction, in which an observer goes more than half way to meet the theorist, is given in an article by Ellison.[23]

We may note, as an aside, that reduction is only one link in the chain, and the absence of agreement between observation and theory may on occasion be traced to faulty reduction. This point also was made by Sherlock Holmes:[24] "I ought to know by this time that when a fact appears to be opposed to a long train of deductions, it invariably proves to be capable of some other interpretation."

The question of bias is very interesting. It is generally recognized that theorists' conclusions are likely to be biased by knowledge of the observations. The reason that great weight is attached to predictions is that these are manifestly free from such bias. It is equally important that facts stated by an observer should be free from bias due to knowledge of theory, but theorists are not so concerned about this possibility that they demand observers to make observations before a theory has been proposed. There is therefore a double standard applied to theorists and observers. However, it would clearly not be possible to require both prediction by the theorists and "pre-observation" by the observers.

It is useful to introduce the term "hard fact" for the case that observations lead to a high probability for one statement of an item and very small probabilities for all alternative statements. The other type of fact may be called a "soft fact." Similarly, we may talk about a "firm conclusion" and a "weak conclusion." We now note that, to get a good test of a theory, we should be able to compare one or more hard facts with one or more firm conclusions. In the case that we are matching a hard fact with a weak conclusion, or a soft fact with a firm conclusion, we are no better off than if we were comparing a soft fact with a weak conclusion. In this case we could say that the strength of the inference is "theory limited" or "observation limited," respectively. The economical use of observational effort and theoretical effort requires a sort of "impedance matching" at the interface!

Sometimes a theory will have one or more adjustable parameters. This possibility may be included in the present formalism by supposing that we are dealing with a continuous sequence of hypotheses H_λ, where λ is a continuous variable. With the notation

$$(H_\lambda \text{ to } H_{\lambda+d\lambda}|X) = (H_\lambda|X)_\lambda\, d\lambda, \quad \text{etc.} \quad , \tag{7.3}$$

Equation (4.10) becomes

$$(H_\lambda|OX)_\lambda = \left[\sum_n \frac{(S_n|H_\lambda X)(S_n|OX)}{\int d\mu (S_n|H_\mu X)(H_\mu|X)_\mu}\right](H_\lambda|X)_\lambda \tag{7.4}$$

and Equation (5.9) becomes

$$(H_\lambda|F_1 \ldots F_A X)_\lambda = \frac{(H_\lambda|F_1 X)_\lambda \ldots (H_\lambda|F_A X)_\lambda \{(H_\lambda|X)_\lambda\}^{-(A-1)}}{\int d\mu (H_\mu|F_1 X)_\mu \ldots (H_\mu|F_A X)_\mu \{(H_\mu|X)_\mu\}^{-(A-1)}} \tag{7.5}$$

This method of determining optimum values of parameters of the theory is very closely related to the "maximum likelihood" method of statistics.[25]

The principal concern of this article has been the problem of establishing a theory as a correct interpretation of a physical phenomenon. When alternative theories can be clearly enumerated, there is the possibility of establishing one of them as correct by proving that the others are incorrect. When alternative theories cannot be clearly enumerated (and this is generally the case), the evidence may point strongly to one of the specified theories, but one must always bear in mind the possibility that further information will come along which will disprove that theory. In this sense, the general situation is that, at any time, there is no "correct" theory of a physical phenomenon--there is only a "front runner."

Although it is instructive, and may sometimes be helpful, to try to specify rules for consistent thinking about scientific theories, any scientist is aware that there are psychological factors as well as rational factors involved in securing acceptance of a theory. It is no doubt the psychological factors which led to the following highly pessimistic observation, attributed to Max Plank: "A new scientific truth does not triumph by convincing the opposition and making them see the light, but rather by the opponents dying off and a new generation growing up to accept it as the truth."

ACKNOWLEDGMENTS

I am grateful to many friends who have listened patiently while I tried to sort out my ideas on this subject. Special thanks for critical and helpful comments are due to M. Parzen, G. Schmidt, G. Tademaru and W. B. Thompson. I wish to thank T. Gold and P. G. Bergmann for permission to quote their unpublished remarks. I also wish to record my debt to E. T. Jaynes who has made a strong case, which I have accepted, for the paradoxical proposition that objective science is based on subjective reasoning.

REFERENCES AND NOTES

1. F. D. Kahn and H. P. Palmer, Quasars (Harvard University Press, Cambridge, Mass., 1967), p. 111.

2. In some professions, this is thought to be good practice. William Murray, Earl of Mansfield, gave the advice "Consider what you think justice requires, and decide accordingly. But never give your reasons; for your judgement will probably be right, but your reasons will certainly be wrong." See Lord Campbell, The Lives of the Chief Justices of England, vol. 3 (James Cockcroft, New York, 1874), p. 481.

3. I. J. Good, Probability and the Weighing of Evidence (Griffin, London, 1950), p. 5.

4. H. Jeffreys, Scientific Inference (Cambridge University Press, 1931), p. 21.

5. G. Polya, Patterns of Plausible Inference (Princeton University Press, Princeton, N. J., 1954), Vol. 2, p. 4.

6. H. Jeffreys, Scientific Inference (Cambridge University Press, 1931), p. 19.

7. I. J. Good, Probability and the Weighing of Evidence (Griffin, London, 1950), p. 16.

8. I. J. Good, Probability and the Weighing of Evidence (Griffin, London, 1950), p. 49.

9. A. S. Eddington, Fundamental Theory (Cambridge University Press, 1949), p. 268.

10. H. N. Russell, Popular Astronomy 22, 275 (1914).

11. H. C. Arp, Handbuch der Physik 51, 75 (1958).

12. H. Jeffreys, Theory of Probability (2nd ed., Oxford University Press, 1948), p. 390.

13. H. Jeffreys, Probability Theory (2nd ed., Oxford University Press, 1948), p. 48.

14. H. Jeffreys, Probability Theory (Oxford University Press, 1948), p. 388.

15. E. T. Jaynes, Probability Theory in Science and Engineering, Colloquium Series in Pure and Applied Science No. 4 (Field Research Laboratory, Socony Mobil Oil Company, Inc., Dallas, 1958), p. 110.

16. There is no up-to-date review article on pulsars. Most articles on this subject have appeared in *Nature*.

17. The neutron star is unusual in that it *began* as a hypothesis.

18. I. J. Good, *Probability and the Weighing of Evidence* (Griffin, London, 1950), p. 63.

19. C. S. McCamy, *Physics Today*, 22, No. 4, 42 (1969).

20. T. Gold has pointed out that "In order for a theory to be useful, it must be possible for it to be proven wrong."

21. Physics Today *22*, No. 6, 63 (1969).

22. This represents a long step towards the end-point (one might say, "the demise") of the pulsar problem: replacing an observational definition by a theoretical definition.

23. M. A. Ellison, Q. J. R. Astr. Soc. 4, 62 (1963).

24. A. Conan Doyle, A Study in Scarlet (The Complete Sherlock Holmes, Doubleday, New York, 1968), p. 49.

25. C. R. Rao, Linear Statistical Inference and its Applications (Wiley, New York, 1965), Ch. 5.

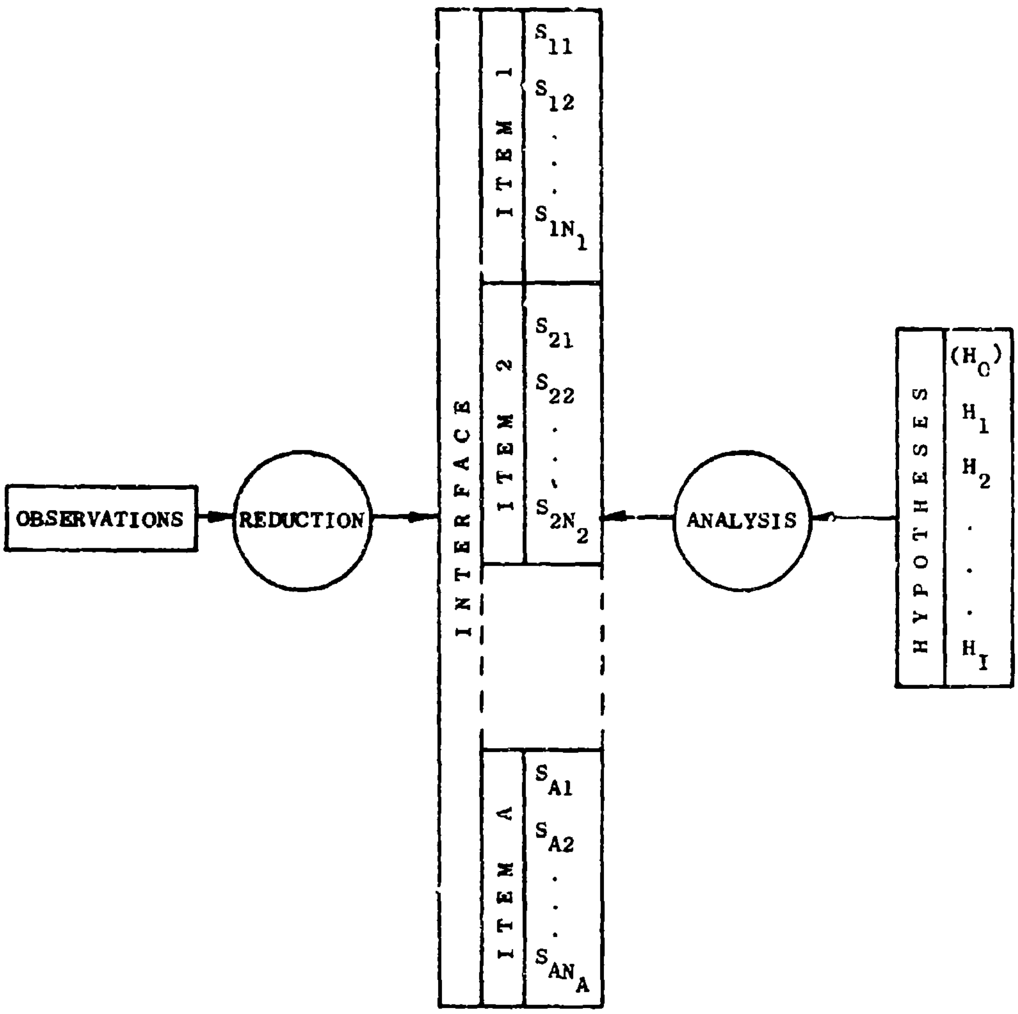

Figure 1.

Schematic Representation of Model used for Evaluation of Relationship between Theory and Observation